=> d bib, kwic 1,3-8,11

L13 ANSWER 1 OF 21 USPATFULL on STN
AN 2003:318650 USPATFULL
TI Computer systems and methods for identifying genes and determining pathways associated with traits
IN Schadt, Eric E., Kirkland, WA, UNITED STATES
Monks, Stephanie A., Seattle, WA, UNITED STATES
PA Rosetta Inpharmatics, LLC, Kirkland, WA, UNITED STATES, 98034 (U.S. corporation)
PI US 2003224394 A1 20031204
AI US 2003-356857 A1 20030203 (10)
PRAI US 2002-353416P 20020201 (60)
US 2002-381437P 20020516 (60)
DT Utility
FS APPLICATION
LREP PENNIE AND EDMONDS, 1155 AVENUE OF THE AMERICAS, NEW YORK, NY, 100362711
CLMN Number of Claims: 163
ECL Exemplary Claim: 1
DRWN 15 Drawing Page(s)
LN.CNT 4513
CLM What is claimed is:

. . . (a) clustering quantitative trait locus data from a plurality of quantitative trait locus analyses to form a quantitative trait locus **interaction** map, wherein each quantitative trait locus analysis in said plurality of quantitative trait locus analyses is performed for a gene. . . . set of genetic markers associated with said species; and (b) identifying a cluster of genes in said quantitative trait locus **interaction** map, thereby identifying members of said biological pathway.

. . . said cluster of genes are those genes that are represented by a gene analysis vector in said quantitative trait locus **interaction** map that shares a correlation coefficient with another gene analysis vector in said in said quantitative trait locus **interaction** map that is higher than 75% of all correlation coefficients computed between gene analysis vectors in said quantitative trait locus **interaction** map.

. . . said cluster of genes are those genes that are represented by a gene analysis vector in said quantitative trait locus **interaction** map that shares a correlation coefficient with another gene analysis vector in said in said quantitative trait locus **interaction** map that is higher than 85% of all correlation coefficients computed between gene analysis vectors in said quantitative trait locus **interaction** map.

. . . said cluster of genes are those genes that are represented by a gene analysis vector in said quantitative trait locus **interaction** map that shares a correlation coefficient with another gene analysis vector in said in said quantitative trait locus **interaction** map that is higher than 95% of all correlation coefficients computed between gene analysis vectors in said quantitative trait locus **interaction** map.

. . . a k-means technique, applying a fuzzy k-means technique, applying a Jarvis-Patrick clustering, applying a self-organizing map technique, or applying a **neural network** technique.

. . . a k-means technique, applying a fuzzy k-means technique, applying a Jarvis-Patrick clustering, applying a self-organizing map technique, or applying a **neural network** technique.

predicting the chemical activity of at least one molecule of interest comprising: a) an input layer consisting of. . . output layer and returning a number between -1 and 1 or another predetermined range; f) a training process for said **neural network** such that said **neural network** can accurately approximate a free energy of **binding** of at least one known training molecule with an output from said output layer; and g) a test process in which a trained **neural network** is used to predict a free energy of **binding** for said at least one molecule of interest; wherein the physicochemical descriptor of said at least one molecule of interest. . . wherein said test process Includes the use of at least one adjuster molecule such that after said training process said **neural network** is used to predict a free energy of **binding** for said at least one adjuster molecule, said at least one adjuster molecule having a known free energy of **binding** and having been excluded from the set of molecules comprising the set of said of at least one known training. . .

2. The **neural network** of claim 1, wherein said **neural network** is able to accurately predict the free energy of **binding** of said at least one adjuster molecule within 10%.

3. A computerized double **neural network** system for predicting the chemical activity of at least one molecule of interest comprising: a) an outer **neural network** further comprising: i) an outer network input layer consisting of at least one neuron where input data is sent as. . . layer and returning a number between -1 and 1 or another predetermined range; vi) a training process for said outer **neural network** such that said **neural network** can accurately approximate a free energy of **binding** of at least one known training molecule with an output from said output layer; b) an inner **neural network** capable of receiving data from said outer **neural network** further comprising: i) an inner network weight matrix where every entry in the form of an input vector is multiplied. . . layer and returning a number between -1 and 1 or another predetermined range; v) a training process for said inner **neural network** such that said **neural network** can accurately approximate a free energy of **binding** of at least one known training molecule with an output from said output layer vi) a test process in which a trained **neural network** is used to predict a free energy of **binding** for said at least one molecule of interest; wherein said inner **neural network** is integrated to function with the data generated from said outer **neural network** such that the rules for said free energy of **binding** learned by said outer **neural network** are utilized by said inner **neural network** to model a quantum object such that said double **neural network** is used to predict the chemical characteristics of said quantum object, said quantum object describing a molecule with improved chemical properties of **binding** relative to said at least one molecule of interest; wherein said outer network output layer is the input layer of said inner **neural network**; wherein said outer network hidden layer includes an error term, said error term being used to calculate the correction terms for said outer network input layer such that the weights and biases of said double **neural network** are optimized; and wherein the physicochemical descriptor of said at least one molecule of interest is the quantum mechanical electrostatic. . .

4. The double **neural network** of claim 3, wherein said test process includes the use of at least one adjuster molecule such that after said outer network training process said **neural network** is used to predict a free energy of **binding** for said at least one adjuster molecule, said at least one adjuster

molecule having a known free energy of **binding** and having been excluded from the set of molecules comprising the set of said of at least one known training. . .
5. The double **neural network** of claim 4, wherein said double **neural network** is able to accurately predict the free energy of **binding** of said at least one adjuster molecule within 10%.

6. The double **neural network** of claim 3, wherein only the weights and biases of said outer network weight matrix are allowed to vary during the training of said double **neural network**.

7. The double **neural network** of claim 3, wherein a bias is added to said outer network hidden layer and said outer network output layer. . .
8. The double **neural network** of claim 3, wherein said outer network hidden layer is composed of 5 hidden layer neurons.

9. The double **neural network** of claim 3, wherein said inner network hidden layer is composed of 5 hidden layer neurons.

10. The double **neural network** of claim 3, wherein said double **neural network** is run through at least 100,000 iterations.

11. The double **neural network** of claim 3, wherein the learning rate of said outer **neural network** is 0.1.

12. The double **neural network** of claim 3, wherein the learning rate of said inner **neural network** is 0.1.

13. The double **neural network** of claim 3, wherein the momentum term of said outer **neural network** is 0.9.

14. The double **neural network** of claim 3, wherein the momentum term of said inner **neural network** is 0.9.

15. The double **neural network** of claim 3, wherein the quantum chemical data sent to said outer network input layer is a vector value derived. . .
16. The double **neural network** of claim 15, wherein said computer is coupled to a display device and there exists a means for presenting the. . .
17. The double **neural network** of claim 3, wherein the process for carrying out the elements of said double **neural network** for predicting the chemical activity of said at least one molecule of interest are contained in a computer, said computer. . .
18. The double **neural network** of claim 17, wherein the chemical characteristics of said quantum object are in the form of a three dimensional representation, said three dimensional representation allowing the identification of the molecular features of said quantum object that said double **neural network** determined could altered to improve the chemical characteristics of said at least one molecule of interest.

19. The double **neural network** of claim 3, wherein said at least one molecule of interest is selected from the group consisting of: a) a. . .

20. The double **neural network** of claim 19, wherein said at least one molecule of interest is an enzyme.

22. The double **neural network** of claim 3, wherein said output value is decreased by at least 1.DELTA.G/RT.

23. The double **neural network** of claim 3, wherein said output value is decreased by 3.DELTA.G/RT.

. . . 24. A computer implemented method for predicting the chemical activity of at least one molecule of interest by using a **neural network** comprising: a) inputting data into an input layer consisting of at least one neuron where input data is sent as. . . generated by said output layer and returning a number between -1 and 1; f) employing a training process for said **neural network** such that said **neural network** can accurately approximate a free energy of **binding** of at least one known training molecule with an output from said output layer; and employing a test process in which a trained **neural network** is used to predict a free energy of **binding** for said at least one molecule of interest wherein the physicochemical descriptor of said at least one molecule of interest. . . wherein said test process includes the use of at least one adjuster molecule such that after said training process said **neural network** is used to predict a free energy of **binding** for said at least one adjuster molecule, said at least one adjuster molecule having a known free energy of **binding** and having been excluded from the set of molecules comprising the set of said of at least one known training. . .

25. The method of claim 24, wherein said **neural network** is able to accurately predict the free energy of **binding** of said at least one adjuster molecule within 10%.

. . . A computer implemented method for predicting the chemical activity of at least one molecule of interest by using a double **neural network** comprising: a) utilizing an outer **neural network** further comprising: i) an outer network input layer consisting of at least one neuron where input data is sent as. . . by said output layer and returning a number between -1 and 1; i) an outer network training process for said **neural network** such that said **neural network** can accurately approximate a free energy of **binding** of at least one known training molecule with an output from said output layer; c) providing an inner **neural network** capable of receiving data from said outer **neural network** further comprising: i) an inner network weight matrix where every entry in the form of an input vector is multiplied. . . by said output layer and returning a number between -1 and 1; v) an inner network training process for said **neural network** such that said **neural network** can accurately approximate a free energy of **binding** of at least one known training molecule with an output from said output layer; vi) a test process in which a trained **neural network** is used to predict a free energy of **binding** for said at least one molecule of interest: d) integrating said inner **neural network** to function with the data generated from said outer **neural network** such that the rules for said free energy of **binding** learned by said outer **neural network** are utilized by said inner **neural network** to model a quantum object such that said double **neural network** is used to predict the chemical characteristics of said quantum object, said quantum object describing a molecule with improved chemical properties of **binding** relative to said at least one molecule of interest; e) constructing said outer network input layer such that said output layer

of said outer **neural network** is the input layer of said inner **neural network**; and f) providing said outer network hidden layer with an error term, said error term being used to calculate the correction terms for said outer network input layer such that the weights and biases of said double **neural network** are optimized; wherein the physicochemical descriptor of said at least one molecule of interest is the quantum mechanical electrostatic potential. . .

27. The double **neural network** of claim 26, wherein said test process includes the use of at least one adjuster molecule such that after said outer network training process said **neural network** is used to predict a free energy of **binding** for said at least one adjuster molecule, said at least one adjuster molecule having a known free energy of **binding** and having been excluded from the set of molecules comprising the set of said of at least one known training. . .

28. The double **neural network** of claim 27, wherein said double **neural network** is able to accurately predict the free energy of **binding** of said at least one adjuster molecule within 10%.

29. The double **neural network** of claim 26, wherein only the weights and biases of said outer network weight matrix are allowed to vary during the training of said double **neural network**.

30. The double **neural network** of claim 26, wherein a bias is added to said outer network hidden layer and said outer network output layer. . .

31. The double **neural network** of claim 26, wherein said outer network hidden layer is composed of 5 hidden layer neurons.

32. The double **neural network** of claim 26, wherein said inner network hidden layer is composed of 5 hidden layer neurons.

33. The double **neural network** of claim 26, wherein said double **neural network** is run through at least 100,000 iterations.

34. The double **neural network** of claim 26, wherein the learning rate of said outer **neural network** is 0.1.

35. The double **neural network** of claim 26, wherein the learning rate of said inner **neural network** is 0.1.

36. The double **neural network** of claim 26, wherein the momentum term of said outer **neural network** is 0.9.

37. The double **neural network** of claim 26, wherein the momentum term of said inner **neural network** is 0.9.

38. The double **neural network** of claim 37, wherein said computer is coupled to a display device and there exists a means for presenting the. . .

39. The double **neural network** of claim 26, wherein the quantum chemical data sent to said outer network input layer is a vector value derived. . .

40. The double **neural network** of claim 26, wherein the process for carrying out the elements of said double **neural network** for predicting the chemical activity of said at least

one molecule of interest are contained in a computer, said computer. . .

41. The double **neural network** of claim 26, wherein said at least one molecule of interest is selected from the group consisting of: a) a. . .

42. The double **neural network** of claim 26, wherein said output value is decreased by at least 1.DELTA.G/RT.

43. The double **neural network** of claim 26, wherein said output value is decreased by 3.DELTA.G/RT.

44. A computerized **neural network** system comprising a **neural network** having a first component trained to recognize **binding** energy for a first set of molecular descriptors based on geometric and/or electrostatic information and for a given **binding** energy returning a second set of the molecular descriptors through a second component of the network.

45. A computerized double **neural network** system comprising a trained **neural network** for predicting **binding** potency for a chemotherapeutic agent with a target molecule, the network having an input layer, and the network being coupled to an output layer of an outer **neural network** comprising one or more layers so that the output of the output layer of the outer **neural network** is the input to the input layer of the inner **neural network**.

47. A computer implemented method comprising providing a **neural network** having a first component trained to recognize **binding** energy for a first set of molecular descriptors based on geometric and/or electrostatic information and for a given **binding** energy returning a second set of the molecular descriptors through a second component of the network.

48. A computer implemented method comprising providing a trained **neural network** for predicting **binding** potency for a chemotherapeutic agent with a target molecule, the network having an input layer, coupling the network to an output layer of an outer **neural network** comprising one or more layers so that the output of the output layer of the outer **neural network** is the input to the input layer of the inner **neural network**.

49. A computer implemented method comprising providing a trained **neural network** for predicting **binding** potency for a chemotherapeutic agent with a target molecule, the network having an input layer coupled to an output layer of an outer **neural network** comprising one or more layers so that the output of the output layer of the outer **neural network** is the input to the input layer of the inner **neural network**, and inputting molecular descriptors based on geometric and/or electrostatic information into the input layer from the coupled outer layer.

51. A computer implemented method of customizing the **binding** features of a molecule of interest comprising: providing a **neural network** comprising a first component trained to recognize **binding** energy for first set of molecular descriptors based on electrostatic and/or geometrical information, and for a given **binding** energy returning a second set of the molecular descriptors through a second component of the network; selecting a molecule of. . .

52. A computer implemented method of determining a set of molecular descriptors: providing a **neural network** comprising an inner network trained to predict **binding** energy of a

molecule of interest with a target molecule using a set of molecular descriptors based on geometric and/or. . . for the molecule of interest the inner network having an input layer coupled to the output layer of an outer **neural network** for inputting molecular descriptors, in the inner **neural network**, setting the **binding** energy for an unknown molecule of interest to a desired level; and determining a set of molecular descriptors for an. . . by computing through the network a set of molecular descriptors that if output from the output layer of the outer **neural network** would yield a **binding** energy within a desired range of a predetermined **binding** energy set for the inner **neural network**.

53. The method of claim 52 wherein the target molecule comprises a **protein** having a **binding** site and the molecular descriptors are for an unknown target molecule that is a potential **binding** agent of the **binding** site and wherein the **binding** energy is set at least slightly above the **binding** energy of a known **binding** agent for the **binding** site.

54. The method of claim 53 wherein the **protein** is an enzyme and the **binding** agent is an inhibitor.

57. The method of claim 52 wherein the **binding** energy level is set to a desired degree higher than the **binding** energy for a known molecule of interest and wherein the method further comprises determining the chemical structure of a molecule. . .

58. The method of claim 57 wherein the determined structure is derived from optimizing **binding** features in a known molecule of interest.

. . . method of claim 57 wherein the determined structure is a modification of a known molecule of interest having a known **binding** energy with the target molecule.

L13 ANSWER 8 OF 21 USPATFULL on STN
AN 2003:173219 USPATFULL
TI High accuracy protein identification
IN Pham, Thang T., Mountain View, CA, UNITED STATES
PI US 2003119063 A1 20030626
AI US 2002-220669 A1 20020903 (10)
WO 2002-US8450 20020318
DT Utility
FS APPLICATION
LREP Jonathan Alan Quine, Quine Intellectual Property Law Group, PO Box 458, Alameda, CA, 94501
CLMN Number of Claims: 59
ECL Exemplary Claim: 1
DRWN 18 Drawing Page(s)
LN.CNT 2554
CAS INDEXING IS AVAILABLE FOR THIS PATENT.
CLM What is claimed is:

1. A method of producing at least one identity candidate for a target **protein** in a sample, comprising: (a) fragmenting **proteins** in a first sample comprising the target **protein** to produce a fragmented sample comprising two or mole peptide fragments of the target **protein**; (b) profiling peptide fragment masses in the fragmented sample by gas phase ion spectrometry under at least two different conditions,. . . by at least one first fractionation technique to produce at least one sub-sample comprising a peptide fragment of the target **protein**, and analyzing one or more sub-samples by the gas phase ion spectrometry to produce at least a second set of. . . mass data; and, (c) querying at least one

database to produce the at least one identity candidate for the target **protein** based upon the first and second sets of peptide fragment mass data.

2. The method of claim 1, wherein the at least one identity candidate identifies the target **protein**.

3. The method of claim 1, wherein the target **protein** comprises at least about 50% by weight of total **protein** in the first sample.

4. The method of claim 1, wherein the target **protein** comprises at least about 50% of the total **protein** molecules in the first sample.

5. The method of claim 1, wherein the **proteins** in the first sample are fragmented enzymatically, chemically, or physically.

6. The method of claim 1, wherein the **proteins** in the first sample are fragmented by one or more proteases.

7. The method of claim 1, comprising producing identity candidates for multiple target **proteins** in the first sample.

11. The method of claim 1, wherein the at least one identity candidate for the target **protein** aids in the diagnosis of one or more pathological conditions.

. . . an initial sample by one or more second fractionation techniques to collect an initial sample fraction that includes the target **protein**, wherein the initial sample fraction is used as the first sample in (a).

. . . of the biomolecules; and (ii) selecting and removing a spot from the array which is suspected of comprising the target **protein**.

. . . a gas phase ion spectrometer, wherein the at least one adsorbent captures one or more peptide fragments from the target **protein** ; (ii) removing non-captured material from the probe, wherein the one or more captured peptide fragments comprise a first sub-sample of. . .

. . . least one support-bound adsorbent, wherein the at least one support-bound adsorbent captures one or more peptide fragments from the target **protein**; (ii) removing non-captured material from the at least one support-bound adsorbent, wherein the one or more captured peptide fragments on. . .

. . . of claim 27, wherein the at least one chromatographic adsorbent comprises one or more of: an electrostatic adsorbent, a hydrophobic **interaction** adsorbent, a hydrophilic **interaction** adsorbent, a salt-promoted **interaction** adsorbent, a reversible covalent **interaction** adsorbent, or a coordinate covalent **interaction** adsorbent.

. . . The method of claims 19, 20, 21, or 22, wherein the at least one adsorbent comprises at least one biomolecular **interaction** adsorbent.

30. The method of claim 29, wherein the at least one biomolecular **interaction** adsorbent comprises one or more of: all affinity adsorbent, a polypeptide, an enzyme, a receptor, or an antibody.

31. The method of claim 29, wherein the at least one biomolecular **interaction** adsorbent specifically captures at least one peptide fragment from the target **protein**.

33. The method of claim 32, wherein the at least one adsorbent comprises

. . . a k-means technique, applying a fuzzy k-means technique, applying a Jarvis-Patrick clustering, applying a self-organizing map technique, or applying a **neural network** technique.

. . . said method further comprises using said cluster of genes in a multivariate analysis to determine whether said genes are genetically **interacting**.

. . . abundance is measured by contacting a gene transcript array with RNA species from said one or more cells, or with **nucleic** acid derived from said RNA species, wherein said gene transcript array comprises a positionally addressable surface with attached **nucleic** acids or **nucleic** acid mimics, said **nucleic** acids or **nucleic** acid mimics capable of hybridizing with said RNA species, or with **nucleic** acid derived from said RNA species.

. . . for clustering quantitative trait locus data from a plurality of quantitative trait locus analyses to form a quantitative trait locus **interaction** map, wherein each quantitative trait locus analysis in said plurality of quantitative trait locus analyses is performed for a gene. . . . genetic markers associated with said species; and (b) instructions for identifying a cluster of genes in said quantitative trait locus **interaction** map, thereby identifying members of said biological pathway.

. . . said cluster of genes are those genes that are represented by a gene analysis vector in said quantitative trait locus **interaction** map that shares a correlation coefficient with another gene analysis vector in said in said quantitative trait locus **interaction** map that is higher than 75% of all correlation coefficients computed between gene analysis vectors in said quantitative trait locus **interaction** map.

. . . said cluster of genes are those genes that are represented by a gene analysis vector in said quantitative trait locus **interaction** map that shares a correlation coefficient with another gene analysis vector in said in said quantitative trait locus **interaction** map that is higher than 85% of all correlation coefficients computed between gene analysis vectors in said quantitative trait locus **interaction** map.

. . . said cluster of genes are those genes that are represented by a gene analysis vector in said quantitative trait locus **interaction** map that shares a correlation coefficient with another gene analysis vector in said in said quantitative trait locus **interaction** map that is higher than 95% of all correlation coefficients computed between gene analysis vectors in said quantitative trait locus **interaction** map.

. . . a k-means technique, applying a fuzzy k-means technique, applying a Jarvis-Patrick clustering, applying a self-organizing map technique, or applying a **neural network** technique.

. . . a k-means technique, applying a fuzzy k-means technique, applying a Jarvis-Patrick clustering, applying a self-organizing map technique, or applying a **neural network** technique.

. . . a k-means technique, applying a fuzzy k-means technique, applying a Jarvis-Patrick clustering, applying a self-organizing map technique, or applying a **neural network** technique.

. . . further comprises instructions for using said cluster of genes in a multivariate analysis to determine whether said genes are genetically

interacting.

. . . for clustering quantitative trait locus data from a plurality of quantitative trait locus analyses to form a quantitative trait locus **interaction** map; wherein (a) instructions for clustering quantitative trait locus data from a plurality of quantitative trait locus analyses to form a quantitative trait locus **interaction** map, wherein each quantitative trait locus analysis in said plurality of quantitative trait locus analyses is performed for a gene. . . genetic markers associated with said species; and (b) instructions for identifying a cluster of genes in said quantitative trait locus **interaction** map, thereby identifying members of said biological pathway.

. . . said cluster of genes are those genes that are represented by a gene analysis vector in said quantitative trait locus **interaction** map that shares a correlation coefficient with another gene analysis vector in said in said quantitative trait locus **interaction** map that is higher than 75% of all correlation coefficients computed between gene analysis vectors in said quantitative trait locus **interaction** map.

. . . said cluster of genes are those genes that are represented by a gene analysis vector in said quantitative trait locus **interaction** map that shares a correlation coefficient with another gene analysis vector in said in said quantitative trait locus **interaction** map that is higher than 85% of all correlation coefficients computed between gene analysis vectors in said quantitative trait locus **interaction** map.

. . . said cluster of genes are those genes that are represented by a gene analysis vector in said quantitative trait locus **interaction** map that shares a correlation coefficient with another gene analysis vector in said in said quantitative trait locus **interaction** map that is higher than 95% of all correlation coefficients computed between gene analysis vectors in said quantitative trait locus **interaction** map.

. . . a k-means technique, applying a fuzzy k-means technique, applying a Jarvis-Patrick clustering, applying a self-organizing map technique, or applying a **neural network** technique.

. . . a k-means technique, applying a fuzzy k-means technique, applying a Jarvis-Patrick clustering, applying a self-organizing map technique, or applying a **neural network** technique.

. . . a k-means technique, applying a fuzzy k-means technique, applying a Jarvis-Patrick clustering, applying a self-organizing map technique, or applying a **neural network** technique.

. . . further comprises instructions for using said cluster of genes in a multivariate analysis to determine whether said genes are genetically **interacting**.

. . . the identification module for clustering said quantitative trait locus data stored in said database to form a quantitative trait locus **interaction** map; wherein a cluster of genes in said quantitative trait locus **interaction** map is identified, thereby identifying members of said biological pathway.

L13 ANSWER 3 OF 21 USPATFULL on STN
AN 2003:270999 USPATFULL
TI Cell-based detection and differentiation of disease states

IN Pressman, Norman J., Glencoe, IL, UNITED STATES
Hirsch, Kenneth S., Redwood City, CA, UNITED STATES
PA Monogen, Inc. (U.S. corporation)
PI US 2003190602 A1 20031009
AI US 2002-241753 A1 20020912 (10)
RLI Continuation-in-part of Ser. No. US 2002-95298, filed on 12 Mar 2002, PENDING
PRAI US 2001-274638P 20010312 (60)
DT Utility
FS APPLICATION
LREP FOLEY AND LARDNER, SUITE 500, 3000 K STREET NW, WASHINGTON, DC, 20007
CLMN Number of Claims: 35
ECL Exemplary Claim: 1
DRWN 10 Drawing Page(s)
LN.CNT 7626
CAS INDEXING IS AVAILABLE FOR THIS PATENT.
CLM What is claimed is:

. . . disease state or discriminating between specific disease states using cell-based diagnosis, comprising a plurality of probes each of which specifically **binds** to a marker associated with a generic or specific disease state, wherein the pattern of **binding** of the component probes of the panel to cells in a cytology specimen is diagnostic of the presence or specific. . .
10. The panel of claim 1, wherein said pattern of **binding** is detected using photonic microscopy.

. . . state or discriminating between disease states in a patient using cell-based diagnosis, comprising: (a) determining the sensitivity and specificity of **binding** of probes each of which specifically **binds** to a member of a library of markers associated with a disease state; and (b) selecting a limited plurality of said probes whose pattern of **binding** is diagnostic for the presence or specific nature of said disease state.

. . . patient known not to be suffering from said disease with each of said probes; (b) measuring the amount of specific **binding** of each probe with its complementary disease marker at loci where said marker is known to be present in cells. . .
. . . The method of claim 13, wherein said selecting comprises one or more of statistical analytical methods, pattern recognition methods and **neural network** analysis.

. . . abnormal cells characteristic of a disease state with a panel according to claim 1; and (b) detecting a pattern of **binding** of said probes that is diagnostic for the presence or specific nature of said disease state.

. . . biochemical biomarker is selected from the group consisting of oncogenes, tumor suppressor genes, tumor antigens, growth factors and receptors, enzymes, **proteins**, prostaglandins and adhesion molecules.

L13 ANSWER 4 OF 21 USPATFULL on STN
AN 2003:266568 USPATFULL
TI Database
IN Swindells, Mark, Easton-on- the- Hill, UNITED KINGDOM
Thornton, Janet, Herts, UNITED KINGDOM
Jones, David, London, UNITED KINGDOM
PI US 2003187587 A1 20031002
AI US 2003-221831 A1 20030204 (10)
WO 2001-GB1105 20010314
PRAI GB 2000-6153 20000314
DT Utility

FS APPLICATION
LREP DARBY & DARBY P.C., P. O. BOX 5257, NEW YORK, NY, 10150-5257
CLMN Number of Claims: 57
ECL Exemplary Claim: 1
DRWN 16 Drawing Page(s)
LN.CNT 3748
CLM What is claimed is:
1) A method of compiling a database containing information relating to the interrelationships between different **protein** and/or **nucleic** acid sequences, said method comprising the steps of: a) integrating data from one or more separate sequence data resources into. . . comparing each query sequence in the combined database with the other sequences represented in the combined database to identify homologous **proteins** or **nucleic** acid sequences; c) compiling the results of the comparisons generated in step b) into a database; and d) annotating the. . .
2) A method of compiling a database containing information relating to the interrelationships between different **protein** sequences, said method comprising the steps of: a) integrating **protein** data from one or more separate sequence data resources and one or more structural data resources into a combined database; b) comparing each query **protein** sequence in the combined database with the other **protein** sequences represented in the combined database to identify homologous **proteins** using, for each query sequence: i) one or more pairwise sequence alignment searches, ii) one or more profile-based sequence alignment. . .
5) A method according to either claim 2 or claim 3, wherein said structural data resource is the **Protein** Data Base (PDB).

. . . A method according to any one of the preceding claims, wherein said integrating step (a) includes the step of scanning **protein** sequences against regular expressions and profiles recorded in a database that contains information relating to annotations of sequence families and. . .
10) A method according to claim 9, wherein **protein** sequences are scanned against regular expressions and profiles in the PROSITE database.

. . . the accessibility potential for each residue to give a total accessibility score; c) summing the pairwise contributions from each residue-residue **interaction** for each of the atom pairs to give a total pairwise energy value; d) inserting the total accessibility score, total pairwise energy value and alignment score into a **neural network** that combines these three values into a single score; and e) comparing this single score to a value calculated for. . .
44) A method according to claim 43, where in said **neural network** is a single-hidden-layer feed forward **neural network**.

46) A database containing information relating to the degree of similarity/interrelationships between different **protein** sequences generated by a method, system or apparatus according to any one of the preceding claims.

47) A database system comprising: a database of **protein** or **nucleic** acid sequence entries containing sequence information, optionally structure information, functional annotation, and information relating to the alignment of each sequence. . .
49) A computer apparatus for compiling a database containing information relating to the similarity between different **proteins**, said apparatus comprising: a processor means comprising: a memory means adapted for storing data relating to amino acid sequences and the relationships shared between different **protein** sequences;

first computer software stored in said computer memory adapted to align said **protein** sequences using one or more pairwise alignment approaches; second computer software stored in said computer memory adapted to align said **protein** sequences using one or more profile-based approaches; third computer software stored in said computer memory adapted to align said **protein** sequences using one or more threading-based approaches.

50) A computer apparatus according to claim 49, wherein said memory means is adapted for storing data relating to: (a) the sequences of a plurality of **proteins** or **nucleic** acids; (b) the structures of a plurality of **proteins**; (c) the predicted alignments of each of said sequences with every other one of said sequences; (d) the predicted alignments. . .
51) A computer apparatus for predicting the biological function of a **protein** comprising: a processor means comprising: a computer memory for storing a specific sequence of amino acid residues; first computer software. . . application programming interface; display means, connected to said processor for visually displaying to a user on command a list of **proteins** with which said specific sequence of amino acid residues is predicted to share a biological function.

52) A computer system for compiling a database containing information relating to the similarity between different **protein** or **nucleic** acid sequences, said system performing the steps of: a) combining sequence data from separate sequence data resources into a composite. . . comparing each query sequence in the composite database with the other sequences represented in the composite database to identify homologous **proteins** or **nucleic** acids using, for each query sequence: i. one or more pairwise sequence alignment searches, ii. one or more profile-based sequence. . .
53) A computer-based system for predicting the biological function of a **protein** comprising the steps of: a) inputting a query sequence of amino acids whose function is to be predicted into a. . .
54) A computer-based system for predicting the biological function of a **protein** comprising the steps of: a) accessing a database according to claim 46 or claim 47, b) inputting a query sequence. . .
55) A computer system for predicting the biological function of a **protein**, comprising: a central processing unit; an input device for inputting requests; an output device; a memory; at least one bus. . . memory storing a module that is configured so that upon receiving a request to predict the biological function of a **protein**, it performs the steps listed in any one of claims 1-45.

56) A computer-based method for predicting the biological function of a **protein**, comprising the steps of: a) accessing the database of claim 46 or 47, at a remote site, b) inputting into. . .
. . . mechanism comprising a module that is configured so that upon receiving a request to predict the biological function of a **protein**, it performs a method as recited in any one of claims 1-45.

L13 ANSWER 5 OF 21 USPATFULL on STN
AN 2003:232019 USPATFULL
TI Methods for predicting functional and structural properties of polypeptides using sequence models
IN Sem, Daniel S., San Diego, CA, UNITED STATES
Baker, Brian, Poway, CA, UNITED STATES
Hansen, Mark R., San Diego, CA, UNITED STATES
PI US 2003162219 A1 20030828
AI US 2001-40895 A1 20011228 (10)
PRAI US 2000-367371P 20001229 (60)
DT Utility

FS APPLICATION
LREP CAMPBELL & FLORES LLP, 4370 LA JOLLA VILLAGE DRIVE, 7TH FLOOR, SAN DIEGO, CA, 92122
CLMN Number of Claims: 32
ECL Exemplary Claim: 1
DRWN 27 Drawing Page(s)
LN.CNT 7513
CAS INDEXING IS AVAILABLE FOR THIS PATENT.
CLM What is claimed is:

1. A method for identifying a polypeptide that **binds** a ligand, comprising: (a) comparing a sequence of a polypeptide to a sequence model for polypeptides that **bind** a ligand, wherein said sequence model comprises representations of amino acids consisting of a subset of amino acids, said subset. . . of amino acids having one or more atom within a selected distance from a bound ligand in said polypeptides that **bind** said ligand; and (b) determining a relationship between said sequence and said sequence model, wherein a correspondence between said sequence and said sequence model identifies said polypeptide as a polypeptide that **binds** said ligand.

2. The method of claim 1, wherein said sequence model comprises a **nucleic** acid sequence.

7. The method of claim 1, wherein one of said sequence models is a **Neural Network** Model.

. . . of amino acids having one or more atom within a selected distance from a bound ligand in said polypeptides that **bind** said ligand.

. . . 9. The method of claim 8, further comprising the steps of: (d) adding a sequence of said identified polypeptide that **binds** said ligand to said set of sequences; and (e) repeating steps (a) through (c) one or more times.

. . . or more atom within a selected distance from a bound conformation of a ligand in a set of polypeptides that **bind** said ligand; and (b) producing a sequence model, amino acids of said sequence model consisting of said subset of amino. . .
12. The method of claim 11, wherein said sequence model comprises a **nucleic** acid sequence.

17. The method of claim 11, wherein one of said sequence models is a **Neural Network** Model.

23. The method of claim 22, wherein said sequence model comprises a **nucleic** acid sequence.

28. The method of claim 22, wherein one of said sequence models is a **Neural Network** Model.

L13 ANSWER 6 OF 21 USPATFULL on STN
AN 2003:227015 USPATFULL
TI Method and apparatus for multivariable analysis of biological measurements
IN Kam, Zvi, Tel Aviv, ISRAEL
PI US 2003158829 A1 20030821
AI US 2002-304551 A1 20021126 (10)
RLI Continuation-in-part of Ser. No. US 1999-345746, filed on 1 Jul 1999, PENDING
DT Utility
FS APPLICATION
LREP FLEIT KAIN GIBBONS GUTMAN & BONGINI, COURVOISIER CENTRE II, SUITE 404,

601 BRICKELL KEY DRIVE, MIAMI, FL, 33131
CLMN Number of Claims: 20
ECL Exemplary Claim: 1
DRWN 12 Drawing Page(s)
LN.CNT 1287
CLM What is claimed is:

1. Apparatus for analyzing multivariable data sets including a plurality of measured variables, said apparatus comprising: a **neural network** capable of receiving signals contained in said data sets and processing said signals according to an artificial intelligence program; and means for obtaining a matrix of weight parameters for said **neural network** and said data sets through a sequence of iterations, starting at random guess, and repeatedly averaging for many initial guesses. . .

. . . of induced and measured variables which characterize stimuli applied to cells and responses of said cells to said stimuli; a **neural network** capable of receiving signals contained in said data sets and to process said signals according to an artificial intelligence program; and means for obtaining a matrix of weight parameters from said **neural network**, said weight parameters allowing identification of fingerprints of complex cellular states.

. . . Apparatus according to claim 3 wherein said external stimulus may be for example a drug, growth factor, hormone, a mutated **proteins** or forced expression of cellular component, and said complex cellular state is starvation, appoptosis, cell differentiation, mitogenicity of proliferating cells. . .

5. Apparatus according to claim 2 for construction of hierarchical architecture of **interaction** network between said components of said biological process by analysis of cell responses to external stimuli, comprising: means for collecting. . . . dependent data sets which includes a plurality of changing variables which characterize responses of said cells to said stimuli; a **neural network** capable of receiving signals contained in said data sets and to process said signals according to an artificial intelligence program; and means for obtaining a matrix of weight parameters from said **neural network**, said weight parameters allowing the construction of hierarchical architecture of said **interaction** network.

7. Apparatus according to claim 1 wherein said **neural network** matrix of weight parameters is represented by a color coded image, in which dominating positive and negative weight parameters are. . .

8. Apparatus according to claim 1 wherein said **neural network** comprises a matrix of weight parameters W.sub.ji which operate on input variables I(k).sub.j, through a monotonic transfer function to generate. . .

9. Process for analyzing multivariable data sets including a plurality of measured variables, said process comprising: providing a **neural network**; applying signals representative of variables contained in said data sets to said **neural network** and processing said data in a sequence of iterations, starting at random guess for said **neural network** matrix of weight parameters, and repeatedly averaging until said matrix of weight parameters converge; and generating from said matrix of. . .

10. Process according to claim 9 wherein said **neural network** comprises a matrix of weight parameters W.sub.ji which operate on input variables I(k).sub.j, through a monotonic transfer function to generate. . .

. . . steps of obtaining a data set comprising a plurality of input/output multivariable vectors representative of a biological process involving many **interacting** multifunctional components in at least one

pathway, establishing a **neural network** comprised of single layer network operators, applying the data set to the **neural network**, training the **neural network** by a training algorithm to implement a transformation by a matrix of weights starting with a first random guess and. . . . . of the same data as the input set, but shifted forward in time to a later time, so that the **neural network** learns to take in data at one time and give out data at the later time.

. . . steps of obtaining a data set comprising a plurality of input multivariable vectors representative of a biological system involving many **interacting** multifunctional components in at least one pathway, that define a complex biological state (or condition), determining a corresponding output vector for each input vector that defines classes for the input vectors, establishing a **neural network** comprised of single layer network operators, applying the data set to the **neural network**, training the **neural network** by a training algorithm to implement a transformation by a matrix of weights starting with a first random guess and. . . plurality of weight matrix solutions are obtained, averaging the weight matrix solutions to obtain an averaged weight matrix, modifying the **neural network** to set its transformation of a matrix of weights to the averaged weight matrix, whereby the modified **neural network** can sort newly presented input vectors into the classes that were defined by the output set of vectors in the. . .

18. Apparatus for processing input vectors representative of a biological process involving many **interacting** multifunctional components in at least one pathway comprising, a **neural network** composed of single layer network operators trained by a training algorithm to implement a transformation by a matrix of weights,. . . of weight matrix solutions are obtained, and then averaging the weight matrix solutions to obtain an averaged weight matrix, said **neural network** having been modified so that the **neural network** is set to its transformation of a matrix of weights to perform its function based on the averaged weight matrix, so that the modified **neural network** will give a predetermined output vector for a predetermined input vector that presents a prescribed pattern of values, and a mechanism for inputting vectors into the modified **neural network**, and to receive output vectors from the modified **neural network**.

L13 ANSWER 7 OF 21 USPATFULL on STN
AN 2003:177147 USPATFULL
TI Method and apparatus for identification and optimization of bioactive compounds using a neural network
IN Braunheim, Benjamin B., 18 E. 105th St., Apt. 12, New York, NY, United States 10029
PA Braunheim, Benjamin B., New York, NY, United States (U.S. individual)
PI US 6587845 B1 20030701
AI US 2000-504407 20000215 (9)
DT Utility
FS GRANTED
EXNAM Primary Examiner: Voeltz, Emanuel Todd; Assistant Examiner: Booker, Kelvin
LREP Ganz, Esq., Bradley M., Ganz Law, PC
CLMN Number of Claims: 59
ECL Exemplary Claim: 1
DRWN 30 Drawing Figure(s); 13 Drawing Page(s)
LN.CNT 1413
CLM What is claimed is:
1. A computerized **neural network** system for

at least one polypeptide that specifically **binds** an immunoglobulin and the method comprises exposing the first or second aliquot to the immunoglobulin, wherein the immunoglobulin specifically **binds** the one or more peptide fragments from the target **protein**, thereby forming a peptide fragment-complex, and contacting the peptide fragment-complex to the at least one adsorbent.

. . . executing an algorithm that determines closeness-of-fit between the computer-readable data and database entries, which entries correspond to masses of identified **proteins** or peptide fragments therefrom, thereby producing the at least one identity candidate for the target **protein** based upon one or more detected peptide fragment masses in the first and second sets of peptide fragment mass data.

. . .

. . . the artificial intelligence algorithm comprises one or more of: a fuzzy logic instruction set, a cluster analysis instruction set, a **neural network**, or a genetic algorithm.

44. A method of producing at least one identity candidate for a target **protein**, comprising: (a) fragmenting **proteins** in a first sample comprising the target **protein** with one or more enzymes to produce a fragmented sample comprising two or more peptide fragments of the target **protein**; (b) profiling peptide fragment masses in the fragmented sample by gas phase ion spectrometry under at least two different conditions,. . . by at least one first fractionation technique to produce at least one sub-sample comprising a peptide fragment of the target **protein**, and analyzing one or more sub-samples by the gas phase ion spectrometry to produce al least a second set of. . . mass data; and, (c) querying at least one database to produce the at least one identity candidate for the target **protein** based upon the first and second sets of peptide fragment mass data.

45. A method of producing at least one identity candidate for a target **protein**, comprising: (a) fragmenting **proteins** in a first sample comprising the target **protein** with trypsin to produce a fragmented sample comprising two or more peptide fragments of the target **protein**; (b) profiling peptide fragment masses in the fragmented sample by surface enhanced desorption/ionization time-of-flight mass spectrometry under at least two. . . of the fragmented sample by affinity chromatography to produce at least one sub-sample comprising a peptide fragment of the target **protein** , and analyzing one or more sub-samples by the surface enhanced desorption/ionization time-of-flight mass spectrometry to produce at least a second. . . mass data; and, (c) querying at least one database to produce the at least one identity candidate for the target **protein** based upon the first and second sets of peptide fragment mass data.

46. A system capable of producing at least one identity candidate for a target **protein** in a sample, comprising: (a) one or more adsorbents capable of capturing peptide fragments in the sample under at least. . . fragment masses in the sets of peptide fragment mass data and database entries, which entries correspond to masses of identified **proteins** or peptide fragments therefrom, thereby producing the at least one identity candidate for the target **protein** based upon the one or more detected peptide fragment masses.

L13 ANSWER 11 OF 21 USPATFULL on STN
AN 2003:112875 USPATFULL
TI Methods and tools for nucleic acid sequence analysis, selection, and generation
IN Hopfinger, Anton J., Lake Forest, IL, UNITED STATES

Riccelli, Peter V., Tinley Park, IL, UNITED STATES
Pancoska, Petr, Evanston, IL, UNITED STATES
Benight, Albert S., Schaumburg, IL, UNITED STATES
PI US 2003077607 A1 20030424
AI US 2002-95923 A1 20020311 (10)
PRAI US 2001-274598P 20010310 (60)
DT Utility
FS APPLICATION
LREP LOUIS MYERS, Fish & Richardson P.C., 225 Franklin Street, Boston, MA, 02110-2804
CLMN Number of Claims: 34
ECL Exemplary Claim: 1
DRWN 14 Drawing Page(s)
LN.CNT 3400
CAS INDEXING IS AVAILABLE FOR THIS PATENT.
CLM What is claimed is:

1. A method of analyzing a **nucleic** acid sequence comprising: constructing a CFD, thereby analyzing a **nucleic** acid sequence.

2. A method of identifying a CFD component associated with a property of a **nucleic** acid sequence or a peptide encoded by the **nucleic** acid, comprising: optionally, providing CFDs for a training set of **nucleic** acid sequences; identifying one or more components of the CFDs; identifying a component, the presence, value, or contribution of which, is correlated, negatively or positively, with a property of the **nucleic** acid or the peptide encoded by a **nucleic** acid, thereby identifying a CFD component associated with a property of a **nucleic** acid sequence or a peptide encoded by the **nucleic** acid.

3. A method of analyzing a **nucleic** acid sequence, comprising: providing a CFD for the **nucleic** acid sequence; identifying one or more components of the CFD; determing if a preselected component, known to be associated with a property of the **nucleic** acid sequence or a peptide encoded by the **nucleic** acid, is present, thereby analyzing the **nucleic** acid sequence.

4. A method of comparing **nucleic** acid sequences, comprising: representing a **nucleic** acid sequence by a mathematical function of the entire sequence context, that depends on the collective characteristics or attributes of. . . . . CFD for each pair of sequences can be represented by three numbers (coefficients) instead of an entire CFD); training a **neural network** or using regression analysis to relate the observed transition temperature and cross-hybridization propensity with the coefficients representative of the CFD of each sequence; optimizing the **neural network** or regression by **interactive** adjustment using algorithms; calculating the predicted CFD from the desired transition temperature and cross hybridization propensity; feeding the desired T.sub.m. . .

11. The method of claim 5, wherein the method is applied to scanning of a **nucleic** acid, e.g., a gene or genome, and finding sequences with most similar and dissimilar segments and includes the following steps:. . .

13. The method of claim 5, wherein the method is used to scan a **nucleic** acid, e.g., a gene or genome sequence, for optimal regions for micro-array applications comprising the following steps. define the T.sub.m. . . define the desired threshold for cross hybridization propensity; define the length of the probes for the microarray; using a trained **neural network** predict the coefficients of the basis CFD's from the desired T.sub.m and cross-hybridization propensity; use the basis CFD's and coefficients. . .

. . . for use in a universal sequence microarray comprising the following steps. (a) generating an Eulerian graph, describing a plurality of **nucleic** acid sequences; (b) partitioning the **nucleic** acid sequences according to a given composition; (c) creating subgraphs that specify how many and what type of the monomeric. . . by their propensity for cross-hybridization by (i) formulating the context functional descriptor of each sequence aligned with itself as a **nucleic** acid duplex at each alignment position and (ii) assigning a number representing the relative thermodynamic stability of the duplex, thereby. . . of the correlation matrix with the deepest minima of the diagonal elements of the correlation matrix, thereby analyzing the potential **interactions** between the **nucleic** acid sequences.

15. The method of claim 5, wherein the method analyzes the potential **interactions** between **nucleic** acid sequences, e.g., sequences described herein, wherein the subgraphs generated in step (c) are listed in a relative manner according. . .

16. A method for analyzing a population of **nucleic** acid sequences comprising: providing a population of **nucleic** acid sequences; providing a CFD for each **nucleic** acid sequence and each **nucleic** sequence of a selected group of complements of the **nucleic** acids of the population; comparing the CFD for each **nucleic** acid sequence and its perfect complement with each of the CFD's for the same **nucleic** acid and each **nucleic** sequence of a selected group of complements of the **nucleic** acids of the population; thereby analyzing a population of **nucleic** acid sequences, e.g., for selecting a subset of the population having a selected degree of cross-hybridization or non cross-hybridization.

21. A method of providing a population of **nucleic** acid sequences comprising: a) providing a value for the length of a **nucleic** acid; b) providing values for the base composition; c) providing a Eulerian representation, of possible sequences which representation can be described by Eulerian graph, d) extracting sequences from the representation, to thereby provide a population of **nucleic** acid sequences.

24. A method of providing a population of **nucleic** acid sequences comprising: a) providing a value for the length of a **nucleic** acid; b) providing values for the base composition; c) providing a representation, sometimes referred to herein as a Eulerian representation,. . . a, b, and c, at least one time; e) extracting sequences from the representations, to thereby provide a population of **nucleic** acid sequences.

26. A method for analyzing **nucleic** acid sequences comprising the steps of: (a) generating an Eulerian graph, or representation thereof, describing a plurality of **nucleic** acid sequences; (b) optionally, partitioning the **nucleic** acid sequences according to a given composition; (c) creating subgraphs that specify how many and what type of the monomeric. . . by their propensity for cross-hybridization by (i) formulating the context functional descriptor of each sequence aligned with itself as a **nucleic** acid duplex at each alignment position and (ii) assigning a number representing the relative thermodynamic stability of the duplex, thereby. . . propensity for hybridization by (i) formulating the context functional descriptor of each sequence aligned with every other sequence as a **nucleic** acid duplex at each alignment position and (ii) assigning a number representing the relative thermodynamic stability of the duplex, thereby. . . of the correlation matrix with the deepest minima of the diagonal elements of the correlation matrix, thereby analyzing the potential **interactions** between the

nucleic acid sequences.

27. A method of and identifying a population of sequences comprising: providing an initial population of **nucleic** acid sequences, e.g., cDNA's; providing, for a first **nucleic** acid sequence of the population, a selected set of oligomers derived from the first **nucleic** acid; providing, for a second and optionally subsequent **nucleic** acid sequence of the population, a selected set of oligomers derived from the second or subsequent **nucleic** acid; providing a T.sub.m, for oligomers produced above and its perfect compliment; selecting subpopulations of the oligomers for which a. . .

28. A method for analyzing a **nucleic** acid sequence, to determine the A T.sub.m involved with introducing a change comprising: providing a **nucleic** acid sequence A and providing a first CFD for the perfect duplex, A, A'; providing a **nucleic** acid sequence B' which is the complement of B and where B differs from A by a change; providing a. . . for the imperfect duplex A, B' by dividing the T.sub.m of A, B' by the correlation coefficient, thereby analyzing a **nucleic** acid sequence.

30. A computer readable file, having a record which includes an element which identifies a **nucleic** acid, and an element which describes the CFD or on or more components thereof.

31. The file of claim 30, wherein the record includes an element which identifies a property of the **nucleic** acid or the peptide it encodes.

32. The file of claim 30, wherein the file includes records for a plurality of **nucleic** acids.

33. A method of analyzing a **nucleic** acid sequence comprising: providing a Eulerian representation of a population of sequences, wherein the population includes at least 10.sup.5 sequences;. . .

34. A set of **nucleic** acids, made or compiled by a method described herein.

35. The set of **nucleic** acids of claim 34, wherein it is an ordered array.